

NELSON-DENNY

READING TEST



Manual for Scoring
and Interpretation

Forms G&H

James E. Brown, Ph. D.

Vivian Vick Fishco, M.S.

Gerald Hanna, Ed. D.



Nelson-Denny Reading Test
Manual for Scoring and Interpretation

Table of Contents

Part 1:	
Purpose and History of the Nelson-Denny Reading Test	1
Part 2:	
Development and Standardization of Forms G and H	2
Part 3:	
Suggested Uses for the Test	8
Part 4:	
Norms	13
References	16
Answer Key	17
Norms Tables:	
Raw Score to Scale Score (SS) to Grade Equivalent (GE) Conversions: Form G	21
Raw Score to Scale Score (SS) to Grade Equivalent (GE) Conversions: Form H	22
Raw Score to Scale Score Conversions: Reading Rate for Forms G and H	23
Raw Score to Scale Score (SS) to Grade Equivalent (GE) Conversions Extended-Time Administration: Form G	24
Raw Score to Scale Score (SS) to Grade Equivalent (GE) Conversions Extended-Time Administration: Form H	25
Form E Raw Scores with Corresponding Form G Scale Scores (SS) and Grade Equivalents (GE): Vocabulary, Comprehension, and Total	26
Form E Raw Scores with Corresponding Form G Scale Scores: Reading Rate	27
Scale Score to Percentile Rank (PR) and Stanine Conversions	28
Percentile Rank (PR) to Normal Curve Equivalent (NCE) Conversions	39
Law Enforcement Academies: Raw Scores with Percentile Ranks (PR) and Stanines: Form G	40
Law Enforcement Academies: Raw Scores with Percentile Ranks (PR) and Stanines: Form H	41

Part 1

Purpose and History of the Nelson-Denny Reading Test

The ability to read well is so important in our culture that it is often the principal cause of success or failure from the first grade of elementary school through college and into professional life. Skillful reading must be mastered and, consequently, measured carefully and fairly.

The primary purpose of the *Nelson-Denny Reading Test*, Forms G and H, is to provide a trustworthy assessment of student ability in three areas of academic achievement: vocabulary, reading comprehension, and reading rate. These important skills are related and interdependent.

From 1929 to 1959, the original forms of the test (Forms A and B), developed by M. S. Nelson and E. C. Denny of Iowa State Teacher's College, served students and teachers as a measure of reading ability and as a guide for instruction. In 1960 Forms A and B were revised by James I. Brown of the University of Minnesota. The original format was retained, except for the addition of a measure of reading rate. The test still could be completed in its entirety in one class period. Forms C and D were completed in 1973 and Forms E and F in 1981.

Forms G and H of the *Nelson-Denny Reading Test* are the latest in a series of revisions of the test that was first administered more than 60 years ago. Content of the test and the statistical data have been periodically updated since the first test in 1929, but the format of this widely used reading survey test remains basically unchanged due to its widespread acceptance.

General Description

The *Nelson-Denny Reading Test (NDRT)* is composed of two subtests, Vocabulary and Comprehension. The Vocabulary section consists of 80 items, each with five answer choices, and has a time limit of 15 minutes. (Previous test forms consisted of 100 items.) The Comprehension section contains seven reading passages and a total of 38 questions, each with five answer choices. (Previous test forms con-

tained eight passages and 36 items.) The time limit for this section is 20 minutes, the first minute being used to determine reading rate. Total administration time for the entire test, including answer sheet preparation, is about 45 minutes.

For Forms G and H of the test, an extended-time administration is available for the first time. This mode of administration allows 24 minutes for the Vocabulary test and 32 minutes for the Comprehension test. Users who exercise the extended-time option should administer the test in two sessions, the first for completing the information required on the answer sheet and taking the Vocabulary test and the second for taking the Comprehension test. (The Reading Rate measure is omitted.)

Test Authors

The senior author of Forms G and H of the *Nelson-Denny Reading Test* is Dr. James I. Brown, Professor Emeritus in the Department of Rhetoric at the University of Minnesota-Twin Cities Campus. Dr. Brown has long been recognized as a principal figure in both test development and college reading circles. He is the author of previous forms of the *Nelson-Denny* and of several college texts for the teaching of reading. Multiple editions of his *Efficient Reading*, *Reading Power*, and *Guide to Effective Reading* attest to the wide use of these texts in colleges throughout the country.

Vivian Vick Fishco is English instructor and Director of the Learning Enhancement Center at Coconino County Community College in Flagstaff, Arizona. Ms. Fishco is a first-time author of the *Nelson-Denny Reading Test* and has collaborated with Dr. Brown on recent editions of *Efficient Reading* and *Reading Power*.

Dr. Gerald S. Hanna is Professor of Education at Kansas State University, Manhattan, Kansas. A notable scholar in the field of tests and measurement, Dr. Hanna was a coauthor of Forms E and F of the *Nelson-Denny Reading Test*.

An Important Note about Test Security

Test materials should be stored in a locked, secure place at all times when not being used.

Part 2

Development and Standardization of Forms G and H

Development of the Test

Development of Forms G and H of the *Nelson-Denny Reading Test* began with a mailing of 1,650 questionnaires to test users in all parts of the country at high school and college levels. The details of the response to this questionnaire are included in the Technical Report. In general, respondents expressed satisfaction with the content and form of the test. Therefore, it was decided that the new edition of the test would have content and form substantially like those of previous editions.

In response to a concern test users expressed about testing time pressures, the Vocabulary subtest was shortened from 100 to 80 items with administration time remaining the same. The Comprehension subtest was changed to include seven rather than eight selections, but with five rather than four questions for each short passage.

The questionnaire indicated that select populations, such as students with English as a second language and returning adults, often need a significantly less-pressured administration of the test. It was decided to develop an extended-time administration for Forms G and H with a special set of norms.

Development of Subtests. After items for Forms G and H were developed, two stages of tryout—informal item tryouts and national item tryouts—preceded the national standardization of the new forms.

For informal item tryouts, test authors worked with large initial pools of items to ensure they would get the high-quality items needed for the final forms. For the Vocabulary test, 766 words were drawn from current, widely used high school and college texts; to ensure maximum relevance, focus was on words that must be known by students in order to cope successfully with school assignments. For the Comprehension test, 571 questions were developed for 32 passages. Passages were drawn from latest editions of humanities, social science, and science texts widely used at the high school and college levels.

From these large initial pools of items, authors were able to choose items for the final forms that improved upon the already impressive statistical properties of their predecessors in Forms E and F.

A complete discussion of the development of the Vocabulary and Comprehension subtests and of the

choice of method used to measure reading rate appears in the Technical Report.

National Item Tryouts. The national item tryouts took place in the fall of 1990. Eight forms of the test were administered to a sample of students in grades 9, 11, 13, and 15. The sample included 7,300 students across these grade levels with approximately 900 students taking each of the eight forms. Males and females were almost equally represented with 54 percent females and 46 percent males. Minority representation consisted of 16 percent black, 4 percent Hispanic, 3 percent Asian/Pacific Islander, and 4 percent Native American. The analyzed data provided the foundation of Forms G and H.

Measures to Ensure Balanced Treatment. In developing Forms G and H, a two-fold attempt was made to address the growing concern about the balanced treatment of minorities and genders in preparation of tests.

To eliminate or minimize bias, the first step was to analyze all 592 vocabulary and comprehension items used in the nationwide fall tryouts. The Mantel-Haenszel Measure of Differential Item Functioning was used; this is perhaps the most widely used procedure for identifying possible bias. Three analyses were performed on each item: black vs. white, Hispanic vs. white, and male vs. female.

This analysis led to the identification of 74 items that were flagged as "possibly biased." Of these items, 49 were eliminated. The remaining 25 items had such high discrimination value that they were retained, so as not to diminish the reliability of the test. However, even these remaining items were chosen so as to obtain the best possible balance between referent and focal groups, thus minimizing favoritism in any one direction.

The second step to ensure fair treatment of minorities and genders was to select a panel of men and women representative of different ethnic groups and ask them to react subjectively to all items and to the reading passages used in the test. While, as might be expected, there was not perfect unanimity, comments of the panelists were helpful in selecting specific passages for use in the final test forms. Out of the 32 passages used in the national tryout, 18 were discarded. The remaining 14 finally selected for Forms G and H were those giving the best possible representation of ethnic and gender groups consistent with acceptable statistical properties.

The members of the balanced treatment review panel:

Dr. Rebecca Bode, Chief Psychologist/Special Services, Cleveland Heights-University Heights City Schools, Ohio

Dr. Frank Ciriza, Researcher/Evaluator, San Diego Unified School District, San Diego, California

Dr. Valeria A. Ford, Director of Student Assessment, District of Columbia Public Schools, Washington, D.C.

Judy A. Hanks, Director of Indian Education, Cass Lake-Bena School District, Cass Lake, Minnesota

Carolyn Lee Hart, Language Arts Teacher, Atlanta Public Schools, Atlanta, Georgia

Pamela Hashimoto, Guidance Counselor, Kenwood Academy, Chicago, Illinois

Fred Hunter and Jeffery Bertrand, English Teachers, Whitney M. Young Magnet High School, Chicago, Illinois

Phat Mekdara, Adviser, Los Angeles Unified School District Division of Adult and Occupational Education Instructional Service Unit, Los Angeles, California; Chairman of Los Angeles County Refugee Communities; Advisory Vice Chair of Association of Korean Education

Dr. Steve Schrankel, Director of Research/Evaluation, Sunnyside Unified School District No. 12, Tucson, Arizona

Dr. Michael W. Strozeski, Director of Research, Garland Independent School District, Garland, Texas

Dr. Joseph H. Suina, Assistant Professor, University of New Mexico, Albuquerque, New Mexico

Dr. Veronica H. Zepeda, Director, Language Development, Roosevelt School District, Phoenix, Arizona

Following the implementation of these procedures to ensure balanced treatment in test development, standardization data were analyzed for bias. This analysis is summarized in the discussion of the test's standardization below.

Standardization of the Test

National Standardization Samples. The national standardization of Forms G and H of the *Nelson-Denny Reading Test* was conducted in September–October 1991 and 1992. Three samples were selected: one from the population of students enrolled in grades 9 through 12 (hereafter referred to as the high school population), one from the two-year college population, and one from the four-year college/university population. In addition, to obtain special norms for law enforcement academies, a sample was

selected from the population of students enrolled in these institutions.

Three stratifying variables were used to guide the selection of participants in the *Nelson-Denny* standardization: geographic region, district enrollment, and the socioeconomic status (SES) of the community. The information related to these variables was obtained from 1980 U. S. Census data. Within each of the four geographic regions, school districts were classified into one of three categories according to their district enrollment. Within each of these region and size categories, schools were further classified into three categories according to the SES index for their communities. The values of this SES index were determined based on two factors: median years of education of the population aged 25 years and older and median family income in thousands of dollars. To derive the SES index value for a given school district, the median value for years of education was multiplied by six and then added to the median income value. This procedure allowed education and income to be weighted approximately equally.

These sampling criteria were used to establish a three-dimensional sampling matrix consisting of 36 cells. After it was determined what fraction of the total sample was desired for each cell in order to achieve proportional representation, sampling within cells was conducted with districts drawn at random.

To secure a more balanced sample, parochial schools were included in the standardization program. The high school sample consisted of nearly 12,000 students from 29 public school districts and 8 parochial schools. A total of 37 public and parochial high schools participated.

The two-year college and four-year college/university samples were selected on the basis of three criteria—four geographic regions of the country, three size categories, and type of institution (public or private). The matrix for two-year colleges consisted of 12 cells with public schools and 4 cells with private schools (one per region). The matrix for four-year colleges/universities consisted of 24 cells. After determining what fraction of the total sample was desired for each kind of institution for each cell, sampling within cells was conducted with colleges drawn at random.

A total of 39 two-year colleges participated in the standardization. The two-year college sample consisted of nearly 5,000 students. From the four-year colleges and universities, 38 institutions participated, the sample totaling over 5,000 students.

To obtain a sample of law enforcement academy students, letters were sent to law enforcement

academies throughout the country announcing the standardization of *Nelson-Denny* Forms G and H and the addition of law enforcement academy norms. The letters resulted in the participation of 16 law enforcement academies and a total of 531 students.

Samples at all levels included representatives of the regions of the country and students of both genders and of various ethnic backgrounds. A detailed description of percentages of students in the standardization by region of the country, gender, and ethnic background is included in the Technical Report.

Participating Schools. Lists of and information about schools and law enforcement academies that participated in the standardization is included in the Technical Report. The publisher and authors wish to express their thanks for the cooperation given by the administrators and teachers in these institutions.

Descriptive Statistics. Tables 1 and 2 provide the descriptive statistics for Forms G and H for beginning and end of year. KR-20 and SEM information are available in the Technical Report, as are the *p*-values and point biserials for each item by group, form, and subtest.

Table 1
Descriptive Data for the Standardization Sample
Beginning of the Year

Grade	High School				2-Yr College		4-Yr College/University			
	9	10	11	12	13	14	13	14	15	16

Form G

Vocab.	Mean	33.27	39.03	42.43	45.58	47.15	51.56	49.40	54.62	58.74	62.72
	SD	13.25	14.37	14.66	14.75	15.16	15.09	15.18	14.59	13.49	11.55
Comp. × 2	Mean	34.45	40.34	42.32	45.22	46.78	50.13	48.50	52.44	55.70	59.07
	SD	15.50	15.96	16.01	15.74	15.61	15.42	15.27	14.84	14.46	12.29
Total (V+2C)	Mean	67.83	79.49	84.50	90.96	93.56	101.65	97.61	106.98	114.55	121.96
	SD	26.91	28.40	29.10	28.46	28.48	28.86	28.94	27.16	26.09	22.02
	N	1799	1714	1483	1271	2023	578	1043	488	584	558

Reading Rate	Mean	211.64	217.04	222.46	226.21	230.21	238.21	234.21	242.50	246.22	250.26
	SD	86.39	86.97	87.14	87.38	87.37	86.80	87.18	85.77	84.88	83.32
	N	1540	1558	1323	1133	1879	509	1009	476	569	544

Form H

Vocab.	Mean	32.87	38.78	42.26	45.48	47.10	51.66	49.42	54.81	59.08	63.21
	SD	13.55	14.74	15.03	15.16	15.59	15.55	15.63	15.03	13.97	11.98
Comp. × 2	Mean	36.68	42.78	44.80	47.73	49.28	52.52	50.94	54.88	57.94	61.28
	SD	16.17	16.32	16.23	15.84	15.61	15.34	15.20	14.58	14.21	11.70
Total (V+2C)	Mean	69.43	81.46	86.58	93.20	95.84	103.97	99.91	109.34	116.92	124.33
	SD	27.90	29.14	29.78	28.99	28.93	29.13	29.30	27.36	26.15	21.84
	N	1651	1620	1327	1113	1797	530	898	447	564	489

Reading Rate	Mean	238.98	244.90	250.84	254.88	259.28	268.09	263.69	272.80	276.93	281.39
	SD	94.53	94.79	94.58	94.92	94.96	94.40	94.78	93.31	92.34	90.62
	N	1423	1516	1188	993	1680	498	852	439	553	478

Table 2
Descriptive Data for the Standardization Sample
End of the Year

Grade	High School				2-Yr College		4-Yr College/University			
	9	10	11	12	13	14	13	14	15	16

Form G

Vocab.	Mean	36.98	41.48	44.59	47.44	50.18	55.74	52.43	58.70	61.21	64.52
	SD	13.26	14.38	14.43	14.75	15.11	14.90	15.12	14.30	13.47	11.46
Comp. × 2	Mean	39.07	42.28	44.25	46.88	50.07	53.31	51.73	55.57	57.28	61.60
	SD	15.54	15.95	15.99	15.47	15.53	15.26	15.19	14.59	14.15	11.94
Total (V+2C)	Mean	74.87	83.58	88.59	94.47	99.41	108.64	103.45	113.94	118.72	126.56
	SD	26.91	27.91	28.60	28.45	28.47	28.78	28.92	27.03	25.74	21.91
	N	1799	1714	1483	1271	2023	578	1043	488	584	558

Reading Rate	Mean	218.68	222.56	227.96	230.44	234.48	242.54	238.46	246.81	251.91	257.65
	SD	86.47	85.75	86.26	87.23	87.32	86.61	87.03	85.54	83.77	83.03
	N	1540	1558	1323	1133	1879	509	1009	476	569	544

Form H

Vocab.	Mean	36.65	41.27	44.46	47.38	50.20	55.91	52.51	58.97	61.57	65.02
	SD	13.56	14.74	14.78	15.16	15.53	15.34	15.54	14.72	13.89	11.83
Comp. × 2	Mean	41.25	44.68	46.69	49.22	52.21	55.40	53.87	57.57	59.39	63.33
	SD	15.77	16.13	16.04	15.37	15.32	14.94	14.89	14.20	13.74	11.38
Total (V+2C)	Mean	76.74	85.70	90.81	96.77	101.74	110.91	105.76	116.21	121.02	128.74
	SD	27.68	28.52	29.14	28.88	28.68	28.74	29.02	26.91	25.60	21.52
	N	1651	1620	1327	1113	1797	530	898	447	564	489

Reading Rate	Mean	246.49	250.72	256.36	259.17	263.46	272.22	267.79	276.86	282.13	288.37
	SD	94.26	93.38	93.47	94.70	94.84	94.34	94.73	93.41	91.42	90.79
	N	1423	1516	1188	993	1680	498	852	439	553	478

Score Comparability and Equating. Studies were conducted to compare Forms G and H of the *Nelson-Denny* to Forms E and F and to show the relationship between the two new forms. A third study was conducted to provide normative data for an extended-time administration of the test. Each of these studies was based on a voluntary sample; the samples are described in the Technical Report.

The Technical Report also gives other details of the old/new and alternate forms reliability studies. The studies showed that the new test correlates well with

the old test and that the alternate forms reliabilities are very high. See Tables 3 and 4 on the following page.

Table 5 on the following page shows the results of the extended-time study, which was conducted using students whose teachers considered them to be disadvantaged readers and/or students for whom their teachers felt more testing time could provide a more accurate assessment. Each student took Form G under extended-time conditions and Form H under normal timing conditions.

Table 3
Means, Standard Deviations, N's, and Correlations for the Old Form (E)/New Form (G) Study

	<i>r</i>	Form E		Form G		N
		Mean	SD	Mean	SD	
Vocabulary	0.86	37.20	17.70	38.74	16.41	1229
Comprehension	0.76	32.06	13.60	38.00	17.02	1229
Total	0.86	69.27	28.97	76.75	31.58	1229
Reading Rate	0.68	243.14	115.77	227.77	99.75	960

Table 4
Means, Standard Deviations, N's, and Correlations for the Alternate Forms Study

	<i>r</i>	Form G		Form H		N
		Mean	SD	Mean	SD	
Vocabulary	0.89	44.82	15.11	44.30	15.75	855
Comprehension	0.81	44.88	16.52	45.84	17.12	855
Total	0.90	89.69	29.89	90.14	30.96	855
Reading Rate	0.68	239.77	90.82	265.88	102.69	703

Table 5
Means, Standard Deviations, N's, and Correlations for the Extended-Time Study

	<i>r</i>	Form G		Form H		N
		Mean	SD	Mean	SD	
Vocabulary	0.83	43.04	11.75	37.60	11.74	215
Comprehension	0.59	50.54	11.40	37.86	12.44	215
Total	0.77	93.57	20.71	75.47	21.84	215
Reading Rate	*	*	*	*	*	*

* No data were collected for Reading Rate

Evidence Addressing Test Fairness. A final measure to address concerns about the balanced treatment of minorities and genders in the preparation of the test was the statistical analysis of bias in the standardization of the test. The procedure used was the Mantel-Haenszel Measure of Differential Item Functioning. There were seven comparisons performed for each form and subtest: male vs. female, white vs. black, white vs. Hispanic, white vs. Asian, white vs. Native American, white vs. Pacific Islander, and white vs. other (the student marked this category if

he/she did not think that one of the other categories was representative of his/her ethnic background).

Each grade level was analyzed separately for each item on each test form. This provided comparisons of ten grade levels for the 118 items on each test form. If an item showed statistically significant differences in performance in more than five grade levels, then the item was considered potentially biased. The only comparison that revealed an item to be statistically biased in more than five grade

levels was the male-female analysis. On Form G, 4 percent of the items were potentially biased in favor of males, and 2.5 percent were potentially biased in favor of females. On Form H, only one item was potentially biased for each male and female

analysis. This indicates that although a few items were statistically biased, the items were balanced across the groups. A more detailed discussion of the Mantel-Haenszel analysis can be found in the Technical Report.

Part 3

Suggested Uses for the Test

Screening Uses

The *Nelson-Denny Reading Test* has several uses as a screening instrument. It may, for example, be used as a general tool for identifying superior students who would profit from placement into an advanced or accelerated program or course. Schoolwide administration of this test should provide useful data for administrators, teachers, counselors, and advisers as they deal with problems of curriculum construction, course organization, and the proper handling of individual differences.

The test is of particular value in identifying students who may need special help developing reading skills in order to take full advantage of their course offerings. For this purpose, the Total score is most useful. It should provide help in determining which students would profit most from placement in a reading clinic, a developmental program, or an accelerated reading class.

An appropriate cutoff point for student placement can be set through experimentation and observation. For a time at the University of Minnesota, for example, the 35th percentile rank was used. Those below that percentile rank were sectioned into a developmental course or encouraged to attend a reading clinic. Those who were above that cutoff point could register for an accelerated class.

The University of Minnesota collected data over a five-year period. Test scores were collected for three groups—those graduating with high distinction, those with distinction, and those on probation. As incoming freshmen, those graduating later with high distinction averaged at the 83rd percentile on the *Nelson-Denny* test, those with distinction at the 68th percentile, and those on probation at the 42nd percentile. At this point, a decision was considered to raise the cutoff score to the 42nd percentile.

A number of educational institutions have successfully used the *Nelson-Denny* as a screening instrument. Following are abstracts of studies that illustrate some of the screening uses for which the *Nelson-Denny* has been found to be valid.

In a study completed in 1993, three California community colleges collected data that indicate a notable relationship between scores on the *Nelson-Denny* and students' classroom performance as shown by course grade. Citrus College, Lassen College, and West Valley College administered the

Nelson-Denny at the beginning of English or reading courses. End-of-course grades were collected and correlated to the *Nelson-Denny* scores. As the following table shows, the correlations between test scores and end-of-course grades were moderate to substantial. Of course, the magnitude of the correlation varies according to the degree to which what is measured on the *NDRT* relates to a particular course and/or the variation in grading standards across classrooms. Note that a test was used to place students in the last three courses shown in the table; because of the restriction of range of test scores, these correlations were corrected to take this into account.

Table 6

**Correlations of End-of-Course Grades and
Nelson-Denny Reading Test Scores in
Three California Community Colleges**

	Vocab.	Comp.	Total
English Composition	0.27	0.30	0.30
Speed and Critical Reading	0.47	0.21	0.37
Reading Fundamentals	0.22	0.33	0.31
Introduction to Reading	0.22	0.45	0.35
Reading Skills Development	0.27	0.64	0.56
Advanced Reading	0.59	0.72	0.72

Lt. S.T. Ragland (1989) reported his experiences in using the *Nelson-Denny* as a screening instrument in a paper discussing the standardizing of Arizona's police pre-employment process. As part of his work with the Arizona Law Enforcement Training Academy, Tucson, Arizona, he administered the test as a pretest during orientation for classes. He then used the results to group the members of the academy classes and to team individuals who had not scored well with individuals who had scored well as tutors or mentors, even attempting to facilitate this tutoring concept by assigning such pairs as roommates in the academy dormitories. He found that students who had not scored well made "immediate and dramatic" improvements. He reported that, after a minimum of 300 recruits had been tested for a statistical base, the academy determined that the grade level 10.9 was critical. He could predict that a police recruit with a *Nelson-Denny* grade level below 10.9 would experience academic difficulties at the academy. The prediction and critical grade level proved accurate with each new recruit class.

Ragland's paper also reported the findings of J.T. Greb, Jr. (1982), who used the *Nelson-Denny* as a predictor of police recruit training success at the Southeast Florida Institute of Criminal Justice. Greb reported that, from a population of 1,373 subjects, the results confirmed the grade 11 equivalent on the *Nelson-Denny* as the critical grade level. Greb compiled six years of test results and correlated academic achievement to *Nelson-Denny* scores. Greb concluded there was "a strong positive correlation" between *Nelson-Denny* test scores and final class average. He also found that recruits with a *Nelson-Denny* grade equivalent of 12 or above tend to achieve higher final class averages and have lower attrition rates.

S. Gudan (1983) reported that a review of the literature indicates the *Nelson-Denny* to be a viable instrument for screening students and predicting their academic success in particular circumstances. In 1981, a study was conducted at Schoolcraft College, Livonia, Michigan, to determine the extent of the relationship between the reading abilities of entering students and their grades in specific classes. Approximately 1,200 freshmen were admitted in the summer of 1981 and were administered the *NDRT* as part of an established orientation program. Those students who took the reading test, actually registered, and received final grades in selected introductory courses were selected for the study. A correlation scattergram showing reading scores and class grades in Basic Biology, General Biology, Introduction to Business, Basic English, English Composition, and Survey of American Government indicated a small positive linear relationship between reading score and class grade. The correlation of reading abilities with the business, English, and political science course grades was statistically significant at the .01 level using the Pearson product correlation, while neither biology course showed strong relationships between reading level and course grade.

T. H. Bers (1982) described an experience of Oakton Community College, Des Plaines, Illinois. In fall 1981, Oakton College students who scored below 35 on the ETS *Test of Written Expression* were not permitted to enroll in C101 (an introductory communications course) until they had successfully completed a developmental communications course or had retaken and passed the exam. To assess this mandatory placement policy, all sections of fall 1980 and fall 1981 C101 courses were compared with respect to grade distribution and course completion rates. Individual student scores on the *Test of Written Expression (TWE)* and the *Nelson-Denny Reading Test (NDRT)* were correlated with measures of academic performance, such as grades in communications courses and grade point average (GPA).

The study revealed that, while *TWE* scores were not significantly correlated to student course grades or overall GPA, *Nelson-Denny* scores were strongly correlated to course grades. It also found that *Nelson-Denny* Total scores (which combine Vocabulary and Comprehension scores) were strongly related to students' GPA's for courses other than communications courses. Based on the findings, recommendations were made to use the *Nelson-Denny* for placement and to track subsequent performance in developmental courses.

Predictive Uses

Closely allied to its screening uses are uses of the *Nelson-Denny* in predicting academic success, a matter of particular concern for administrators, counselors, and teachers at both the high school and college levels.

In advising a student about possible courses, an adviser or a counselor should be keenly aware of the student's reading level, since some educational goals or majors demand much more reading ability than others.

A number of schools have conducted studies that show the *Nelson-Denny* can be used as a predictor of student success. Some of their experiences are summarized in the following abstracts.

H. H. Allen (1991) found a statistically significant positive relationship between reading achievement of college freshmen as measured by the total score of the Texas Academic Skills Program (TASP) and as measured by the Total score of the *Nelson-Denny Reading Test*.

W. M. Whitton (1990) investigated whether the *Nelson-Denny* was an effective instrument to predict academic performance for recruits enrolled in the Basic Law Enforcement Training Course of the University of Houston Downtown Regional Law Enforcement Academy and on the post-academy State of Texas Licensing Examination. Analysis of the data collected indicated positive correlations between all three *Nelson-Denny Reading Test* scores (Vocabulary, Comprehension, and Total) and the recruits' scores on the mid-term and final examinations, overall class averages, and scores on the State of Texas Licensing Examination.

In a study conducted by R.C. Feldt (1988), prediction of performance in an introductory psychology course was examined for 65 first-year undergraduates. Individuals completed the Comprehension subtest of the *Nelson-Denny* and were asked to study a 956-word passage, report their reading strategies, and respond to 18 test questions on the passage. Predictors of performance included *Nelson-Denny* Comprehension score, reading strategy,

study time, retention interval, time to complete the test, and test score. Percentage of total points earned in the course for the semester was the performance criterion. The best set of predictors included Comprehension score, test score, and retention interval.

M. Webb (1984) conducted a study at Florissant Valley Community College (FVCC) to compare three standardized tests, among them the *Nelson-Denny*, for use in assessing student reading skills. The study population of 402 students included 216 students who agreed to participate and 186 students enrolled in freshman English classes selected for testing. At the end of the fall 1982 semester, course grades were obtained for the tested students. Predictive validity of the three tests was determined by computing Pearson product-moment correlation coefficients. A step-wise regression analysis was performed to determine the most effective combination of variables for predicting academic performance. Study findings revealed that all three tests had adequate predictive validity.

C. E. Heerman and K. N. Seltzer (1983) concluded from their research that the *Nelson-Denny* provides valid predictions of student academic aptitude and is a reliable measure of reading attitude.

K. Hawes (1982) reported that correlations between the *Nelson-Denny Reading Test*, Form E, and the *American College Testing Program* composite scores of 168 undergraduates were similar to those reported in the *NDRT* manual. This supports the *Nelson-Denny's* authors' suggestion that the *NDRT* can be used to predict academic success.

P. H. Wood (1982) conducted a study to determine the difficulty level, the internal consistency, and the usefulness of the *Nelson-Denny* as a predictor of college freshmen grades. He found that adding the *NDRT* to a prediction equation already containing high school grade point average improved the prediction.

The work of T. D. Erwin (1981) provided further evidence of predictive validity of the *Nelson-Denny* for college students. This study argued that because the *Nelson-Denny* is more convenient and economical to administer, it can serve as a feasible alternative to the *SAT* when a quick verbal measure is needed in the college setting.

Erwin and J. L. Millikin (1980) found the *Scholastic Aptitude Test* verbal, reading comprehension, and vocabulary scores similar to the *Nelson-Denny* Total, Comprehension, and Vocabulary scores. Correlation coefficients developed from these comparisons,

when viewed as validity coefficients, were moderately high.

In a study conducted by J. R. Gerow and D. P. Murphy (1980), the median correlation of *Nelson-Denny* scores with first unit examination and semester point total was 0.56. When standardized aptitude scores are not available, they concluded, the *Nelson-Denny* would be a useful predictor of performance in an introductory psychology class.

Diagnostic Uses

Although the *Nelson-Denny Reading Test* is intended primarily as a survey instrument, it can also serve a limited diagnostic function, contributing insights of importance to teachers in facilitating student progress.

Revised Forms A and B first introduced a three-dimensional diagnostic pattern of value in determining individual reading strengths and weaknesses with scores in the key areas of vocabulary, comprehension, and reading rate. This three-fold pattern has remained a constant in all subsequent forms of the test. This was intended to help teachers cope more effectively with major reading problem areas.

For the revision of the test that has resulted in Forms G and H, an extensive review was undertaken of developmental reading textbooks currently in use and/or offered by major publishers. This review confirmed the validity and the currency of the test's approach of examining these three key areas of vocabulary, comprehension, and reading rate.

Examining Student Scores

If a student's percentile ranks tend to be approximately the same in all three areas, the teacher might assume that no one of these areas is a particular problem. Attention to all three would seem in order in planning instruction.

More often than not, however, a student's test profile will show one area well above or below the others. Generally speaking, the greater the deviation between subtests, the greater the need for follow-up instruction in an area of low achievement. For example, suppose a student's percentile ranks are as follows: Vocabulary, 15th percentile; Comprehension, 45th percentile; and Reading Rate, 50th percentile. These scores tend to point to vocabulary as a major problem. In spite of the low percentile rank in vocabulary, it is encouraging to note that the student is making better than expected use of present word power. (An expected use of word power would indicate that the student's score on the Comprehension subtest would likely have been in the 15th percentile, rather than in the 45th percentile.) The teacher should probably stress vocabulary building for this student. It would seem a safe assumption

that special efforts in improving vocabulary would mean further improvements in comprehension, for there is no reason to believe that the student will not continue to make good use of a more fully developed vocabulary. Intensive work on vocabulary should bring measurable progress in both vocabulary and comprehension in a subsequent testing.

Another student might have a higher percentile rank in vocabulary than in comprehension: Vocabulary, 65th percentile; Comprehension, 45th percentile; and Reading Rate, 20th percentile. These percentile ranks suggest that the student already has the word power to achieve the 65th percentile in comprehension, but is not using that word power as expected. The difficulty may be motivational, or it may be due to a deficiency in background knowledge or a variety of other causes.

The following "rule of thumb" may be helpful to a teacher in looking at a student's strengths and weaknesses on the *Nelson-Denny*. Look at the percentile rank (PR) for the Total score. If it is between 25 and 75, then Vocabulary, Comprehension, and Reading Rate scores must be 18 percentile points above (or below) the Total score PR to be considered high (or low) for that student. If the PR for the Total score is between 76 and 87 (or 13 and 24), then subtest scores must be 12 or more percentile points above (or below) the Total score PR to be considered high (or low). If the Total score PR is 88 or greater (or 12 or less), then subtest scores must be 6 or more percentile points greater (or less) to be considered high (or low).

Students' reading rates can also provide clues to their reading difficulties. When the reading rate score is substantially below the other scores, particularly when it is in the bottom quartile, it suggests the strong possibility that the student has habits of regressing, vocalizing, or word-by-word reading. Improvement of reading rate should, of itself, result in some improvement in both the Vocabulary and Comprehension subtests if no other factors are operating to cause a student's reading problems.

A comparison of the reading rate percentile with the comprehension percentile will focus attention on individual differences and problems, as well as on identification of reading patterns. For example, one student may read rapidly but inaccurately, another rapidly and accurately, a third slowly and inaccurately, and a fourth slowly and accurately. Knowledge of a student's reading rate coupled with the Comprehension score can help one decide the amount of attention to devote to reading speed in the context of overall improvement of reading effectiveness.

Using Item Classifications to Examine Student Scores

The items of the *Nelson-Denny* Comprehension subtests were originally constructed to fit into a classification of varying degrees of literalness and interpretation. In such a classification, items are somewhat arbitrarily designated literal or interpretive, but the thirty-eight items on each form can be divided into a group of nineteen largely interpretive and nineteen primarily literal items. Table 7 indicates which test items fall into each of the two categories.

Comprehension passages were chosen from a variety of subject matter fields so that the test would not favor students strong in any one subject matter area. Table 8 on the following page presents the spectrum of areas represented by the passages and the items from each passage.

The major diagnostic use most teachers will make of the *Nelson-Denny* will be to focus on differences among Vocabulary, Comprehension, and Reading Rate scores. Grouping test items by different categories, as in Tables 7 and 8, is helpful in surveying the coverage of the test. It is less useful in working with individual students, however, due to measurement error with small numbers of items in the categories, high intercorrelations among categories, and the necessity of a teacher's doing extra work with this kind of analysis. Yet some teachers may wish to use such analysis with atypical students who present reading problems that are particularly vexing.

Table 7
Comprehension Item Numbers by Category

Literal Type		Interpretive Type	
Form G		Form G	
1, 2, 3, 4, 9, 10, 14, 15, 16, 19, 20, 21, 24, 25, 26, 29, 31, 34, 35		5, 6, 7, 8, 11, 12, 13, 17, 18, 22, 23, 27, 28, 30, 32, 33, 36, 37, 38	
Form H		Form H	
1, 2, 3, 4, 9, 10, 11, 14, 16, 19, 20, 21, 24, 25, 28, 29, 33, 34, 35		5, 6, 7, 8, 12, 13, 15, 17, 18, 22, 23, 26, 27, 30, 31, 32, 36, 37, 38	

Table 8
Comprehension Item Numbers by Subject Category

Form G		Form H	
I. Humanities			
Passage	Items	Passage	Items
I	1, 2, 3, 4, 5, 6, 7, 8	I	1, 2, 3, 4, 5, 6, 7, 8
*II	9, 10, 11, 12, 13	*III	14, 15, 16, 17, 18
II. Social Sciences			
**IV	19, 20, 21, 22, 23	**IV	19, 20, 21, 22, 23
**VII	34, 35, 36, 37, 38	**VII	34, 35, 36, 37, 38
III. Science			
*III	14, 15, 16, 17, 18	*II	9, 10, 11, 12, 13
**V	24, 25, 26, 27, 28	**V	24, 25, 26, 27, 28
*VI	29, 30, 31, 32, 33	*VI	29, 30, 31, 32, 33

* passages drawn directly from current widely used high school texts

** passages drawn directly from current widely used college texts

Part 4

Norms

Derived Scores

This section provides information important in interpreting and using results of the *Nelson-Denny Reading Test*. A raw number or percentage of test items answered correctly cannot be interpreted without highly specific information about the content of a test. Most uses of standardized achievement tests require comparison of an examinee's performance with the performance of relevant reference groups. A variety of derived scores have been developed to enable users of the *Nelson-Denny* to make these comparisons.

Reference Groups

Norm-referenced test interpretations can be based on data from various kinds of reference groups; many testing purposes are best served by reference groups that are broadly representative of the national population. The use of a national reference group permits a comparison of individuals with a representative national cross section of their counterparts. National reference groups are well suited to the instructional uses of the *Nelson-Denny*. All normative data provided in this manual are based on national reference groups.

As described in Part 2 of this manual, empirical *Nelson-Denny* normative data were gathered for each of grades 9 through 12 at the high school level, for grades 13 and 14 at the two-year college level, and for each of grades 13 through 16 at the four-year college/university level. After smoothing the resulting distributions of scores, these fall ("Beginning of the Year") data were interpolated and extrapolated to provide spring ("End of the Year") data. In addition, normative data were gathered for an extended-time administration of the test and for law enforcement academies. Because of the particular needs of law enforcement academies, these norms are not presented as fall and spring norms but are combined.

Scores Provided

Derived scores may be obtained for each of the *Nelson-Denny* subtests (Vocabulary, Comprehension, Reading Rate) as well as for the Total—Vocabulary plus two times Comprehension. (The reason for double-weighting Comprehension is that its standard deviation for this relatively short test—38 items—is about half as much as Vocabulary—80 items.)

This section outlines the characteristics of the *Nelson-Denny* scores. Raw scores are obtained by following the directions for scoring the test. Users generally are advised not to attempt interpretations by means of raw scores. To render the examinee's performance comparable with that of one or more reference groups, raw scores are converted into derived scores.

The only subtest with which many users might be tempted to use raw scores is Reading Rate. There is a tradition of reporting reading rate in terms of words-per-minute scores. However, rate depends directly on the passage being read and upon the purpose for which it is being read. It is inappropriate to generalize a words-per-minute score of any reading rate test to other passages or circumstances. Therefore, users of the Reading Rate subtest are advised to use one of the derived scores provided in interpreting results of this subtest as well as in using all other scores of the *Nelson-Denny*.

Five kinds of derived scores are supplied in this manual. *Percentile ranks* are best suited for the evaluative and instructional interpretations that instructors and other educators frequently make. *Stanines* are provided for users who do not require fine discriminations among students. *Normal curve equivalent scores* (NCES's) are useful for computing averages and making comparisons. Each of these derived scores is based on a national reference group of students in the same group as the examinees whose scores are being interpreted. For users desiring to measure growth or to perform certain kinds of statistical calculations with derived scores, *scale scores* are provided. Finally *grade equivalents* are reported.

Percentile Ranks. A mid-interval percentile rank reports the percentage of a reference group that an examinee surpassed plus one-half of the percentage that the examinee tied. This kind of score is easily interpreted to students, instructors, and parents. Percentile ranks are recommended for most interpretive and instructional uses of the *Nelson-Denny*.

In interpreting these mid-interval percentile ranks, care should be taken to ensure that it is understood that a score refers to the percentage of people in a reference group who obtained scores lower than the one being considered, not to the percentage of questions answered correctly. A person's percentile rank of 31 means that his/her score was equal to or higher than 31 percent of his/her grade peers tested at the same time of year.

Stanines. Stanine scores are based on a nine-interval normalized standard score scale ranging from a low of 1 to a high of 9. Stanines in a normal distribution have a mean of 5 and a standard deviation of 2. The meaning of stanines is best understood through their relationship to percentile ranks. This relationship is shown here.

Percentile Ranks	Stanine
96-99	9
89-95	8
77-88	7
60-76	6
40-59	5
23-39	4
11-22	3
4-10	2
1-3	1

Stanines enjoy certain advantages of convenience and ease of statistical manipulation. On the other hand, they are not as refined as the other kinds of derived scores provided in this manual.

Normal Curve Equivalents. The normal curve equivalent score is a standard score with a mean of 50 and a standard deviation of 21.06. Like the percentile rank, this score has a range from 1 to 99, but normal curve equivalents do represent *equal units*. They can be used to compute averages and to make comparisons. The following chart illustrates the relationships of stanines and percentile ranks to normal curve equivalent scores.

NCE-PR Relationships		NCE-Stanine Relationships	
NCE	PR	NCE	Stanine
99	99	86 up	9
90	97	77-85	8
80	92	66-76	7
70	83	56-65	6
60	68-69	45-55	5
50	50	35-44	4
40	31-32	24-34	3
30	17	15-23	2
20	8	1-14	1
10	3		
1	1		

Scale Scores. Scale scores are recommended for assessing growth over a period of time. These scores possess the equal-interval characteristic that is important for certain statistical calculations.

Scale scores are based upon the pooled standardization samples from grades 10, 11, and 12, both two-year college classes, and both lower-division classes in the four-year institutions. The unweighted pooled distribution from these seven reference groups has been converted into a system of normalized scale scores for which the mean is set at 200 and the standard deviation at 25.

Grade Equivalents. *Nelson-Denny* grade equivalent scores report the grade at which an examinee's test performance would be at the median. For example, if Terry, an eleventh grader, received a grade equivalent of 9.2, her performance on the test she took was as high as the average of students who are in the second month of the ninth grade.

Nelson-Denny grade equivalent scores are based upon the regular standardization sample for grades 9 through 12, upon the unweighted combined two-year college and four-year institution standardization samples for grades 13 and 14, and upon the four-year college/university standardization samples for grades 15 and 16. Since there is a great deal of attrition between the freshman year of high school and the senior year of college, grade equivalent scores are not based upon comparable samples at all grade levels; attrition creates an illusion of more growth than actually occurs. Grade equivalents extending upward to 18.9 and downward to 4.1 were obtained by extrapolation.

Grade equivalents suffer from several serious limitations that may result in their misuse and misinterpretation. Following are three of the characteristics of grade equivalents that should be kept in mind by persons who consider using them with *any* test.

- Grade equivalents do not necessarily indicate the grade placement at which the student could adequately perform. Rather, they indicate the grade placement at which the performance would be at the average *on the test taken*. For example, the grade equivalent of 16.3 for Maria, who is in the eleventh grade, does not necessarily indicate that she could satisfactorily handle upper-division university textbooks.
- Grade equivalents do not provide a uniform unit of measurement. Like all age and grade scales, they cease to be meaningful beyond the growing years. The rate of growth in reading slows down before and during the range of grades for which the *Nelson-Denny* is designed. Therefore, one should not expect the difference between grade equivalents of, say, 4.5 and 5.5 to be the same as the difference between 15.5 and 16.5. The former would be much greater. It is impossible to represent the status of an above-

average university senior by means of these grade equivalents.

- Grade equivalents do not necessarily have the same variability among tests that measure different content. They therefore are particularly hazardous in profile interpretations.

Grade equivalents should not be used to report an examinee's within-grade standing or relative performance on different tests or subtests. They could be used for measuring growth, but the authors believe that the scale scores provided are far superior for this purpose.

Use of Normative Tables

To translate raw score into scores on the norms tables, use the instructions that follow for each type of table.

Scale Scores and Grade Equivalents. Tables on pages 21–23 provide the normalized scale scores and grade equivalents corresponding to raw scores for each form of the *Nelson-Denny*.

Extended-Time Norms. Scale scores and grade equivalents for students who participate in the extended-time administration of the test are provided on pages 24–25. If other scores are required, use the scale score to obtain a percentile rank or stanine.

Old Form/New Form Data. Data on pages 26–27 translate raw scores on Form E to corresponding scale scores and grade equivalents on Form G. If other scores are required, use the scale score to obtain a percentile rank or stanine.

Percentile Ranks and Stanines. To obtain a grade-based percentile rank or stanine, one must first convert a student's raw score to a scale score

using the tables on pages 21–23. Then one must turn to the tables listed on page 28 to find that scale score in the table for the appropriate reference group.

This manual presents tables for many different reference groups. Scores of examinees tested during the fall term of a school year are obtained from the fall ("Beginning of the Year") table appropriate for the student's grade or class, while scores of students examined during the spring term are secured from the spring ("End of the Year") table corresponding to the student's class or grade.

The figures displayed in the tables under "Comprehension" represent raw scores on the Comprehension test *that have been doubled*. To use the tables, one who uses the Self-Scorable Answer Sheet must follow the directions on the back of that sheet and double the actual raw score on Comprehension obtained by the student. *Those who hand-score other answer sheets must also remember to double this raw score.*

The tables listed on page 28 show percentile ranks and stanines for both forms of the test.

Normal Curve Equivalents. A table on page 39 provides normal curve equivalent scores (NCE's) corresponding to percentile ranks for both forms.

Law Enforcement Academy Norms. Forms G and H mark the first time that normative data for the *Nelson-Denny* have been collected specifically for law enforcement academies. These norms can be found on pages 40–41. (Note that law enforcement academy tables do not present both fall and spring norms.)

Law enforcement academy tables present only percentile ranks and stanines. For scale scores and grade equivalents, use the tables on pages 21–23.

References

- Allen, H. H. (1991). *A correlational study of achievement as measured by the total score on the Texas Academic Skills Program (TASP) Reading Test and achievement as measured by the total score on the Nelson-Denny Reading Test*. Doctoral dissertation, University of Houston.
- Bers, T. H. (1982). *Assessment of mandatory placement in communications: Fall 1981*. Office of Institutional Research, Oakton Community College, Des Plaines, IL.
- Erwin, T. D. (1981). The Nelson-Denny Reading Test as a predictor of college grades. *Reading Psychology*, 2, 158-164.
- Erwin, T. D. and Millikin, J. L. (1980). The relationship of the Nelson-Denny Reading Test to the Scholastic Aptitude Verbal Tests. *Measurement and Evaluation in Guidance*, 13, 169-171.
- Feldt, R. C. (1988). Predicting academic performance: Nelson-Denny Reading Test and measures of college students' study of expository prose. *Psychological Reports*, 63, 579-582.
- Gerow, J. R. and Murphy, D. P. (1980). The validity of the Nelson-Denny Reading Test as a predictor of performance in introductory psychology. *Educational and Psychological Measurement*, 40, 553-556.
- Greb, J.T., Jr. (1982). *The Nelson-Denny Reading Test as a predictor of police recruit training success and the impact of basic reading skill levels over a six-year period*. Doctoral dissertation, Florida Atlantic University.
- Gudan, S. (1983). *The Nelson-Denny Reading Test as a predictor of academic success in selected classes in a specific community college*. Schoolcraft College, Livonia, MI.
- Hawes, K. (1982). Correlations of scores on the Nelson-Denny Reading Test, Form E, and measures of academic aptitude, academic achievement, readers' attitude, and students' self-confidence. *Psychological Reports*, 51, 229-230.
- Heerman, C. E., and Seltzer, K. N. (1983). Correlating reading and academic variables with the Nelson-Denny Test. *Reading Improvement*, 20, 6-66.
- Ragland, S. T. (1989). *Nelson-Denny experiment: standardizing Arizona's police pre-employment process*. Arizona Law Enforcement Training Academy, Tucson, AZ.
- Webb, M. (1984). *A comparison of the ISRT, NDRT, and SDRT for use in assessing FVCC student reading skills*. Saint Louis Community College at Florissant Valley, MO.
- Whitton, W. M. (1990). *The Nelson-Denny Reading Test as a predictor of academic performance of police recruits and the impact of nine related variables on recruit academic performance*. Doctoral dissertation, The Union Institute.
- Wood, P. H. (1982). The Nelson-Denny Reading Test as a predictor of college freshman grades. *Educational and Psychological Measurement*, 42, 575.